



## Internet: más allá de la superficie

### Internet: Beyond The Surface

### Internet: para além da superfície

Diego Andrés Tovar Cardozo<sup>1</sup>

Luis Miguel Pimienta Acero<sup>2</sup>

Edna Rocío Ramírez Saavedra<sup>3</sup>

**Fecha de recepción:** junio 2014

**Fecha de aceptación:** noviembre 2014

**Para citar este artículo:** Tovar, D. Pimienta, L. y Ramírez, E. (2015). Internet: más allá de la superficie. *Revista Científica*, 21, 81-90. **Doi:** [10.14483/udistrital.jour.RC.2015.21.a08](https://doi.org/10.14483/udistrital.jour.RC.2015.21.a08)

### Resumen

El presente artículo expone un enfoque técnico de la Internet que da a conocer algunas características de su estructura, la cual muestra dos sectores distintos en donde se puede encontrar la información, es decir, esta se encuentra distribuida en niveles o capas de manera parcial o restringida, integrada o desagregada, dependiendo de lo que el usuario requiera. La forma correcta para acceder a la información es saber a dónde buscarla, pues la web profunda es información no visible para el usuario común que no tiene la habilidad de búsqueda, por lo cual se hace necesario verificar el tipo, la calidad y veracidad de la información a obtener antes de proceder a usar aquella encontrada con los buscadores tradicionales. Acceder a Internet sin saber que existe una falsa cortina utilizada por la mayoría de los buscadores convencionales es limitar las capacidades de

navegar en un océano de información valiosa; en ello radica la importancia de conocer la web desde lo superficial (*surface web*) a lo profundo (*deep web*).

**Palabras Clave:** web superficial, TCP/IP, URL, WWW, motores de búsqueda, HTML, web profunda.

### Abstract

This paper presents a technical approach to the Internet that discloses some features of its structure, which shows two distinct sectors where you can find information, ie, this is distributed in levels or layers partially or restricted, integrated or broken, all depending on what the user really requires. The correct way to access the required information is in knowing where to look, deep web is not visible to the average user who does not have the ability to search information, which is necessary to verify the type, quality

<sup>1</sup> Universidad de la Amazonia, Florencia Caquetá. Colombia. Contacto: [diegoandrestovar@gmail.com](mailto:diegoandrestovar@gmail.com)

<sup>2</sup> Universidad de la Amazonia, Florencia Caquetá. Colombia. Contacto: [luismil\\_p@hotmail.com](mailto:luismil_p@hotmail.com)

<sup>3</sup> Universidad de la Amazonia, Florencia Caquetá. Colombia. Contacto: [ramsa0223@hotmail.com](mailto:ramsa0223@hotmail.com)

and veracity of the information to be obtained prior to using the information found in traditional search engines. Access the Internet without knowing that there is a false curtain used by most conventional search engine, is to limit the capabilities of navigating an ocean of valuable information, therein lies the importance of knowing the web from the surface (*Surface Web*) to deep (*Deep Web*).

**Keywords:** Surface Web, TCP/IP, URL, WWW, Search Engines, HTML, Deep Web.

### Resumo

Este artigo apresenta uma abordagem técnica para a Internet que revela algumas características de sua estrutura, o que mostra dois setores distintos, onde você pode encontrar informações, ou seja, este é distribuído em camadas ou níveis parcial ou restrito, integrado ou quebrado, tudo dependendo do que o usuário requer realmente. A maneira correta de acessar as informações necessárias está em saber para onde olhar, deep web não é visível para o usuário médio que não tem a capacidade de buscar informações, o que é necessário para verificar o tipo, qualidade e veracidade da informação a ser obtida antes de usar as informações encontradas nos motores de busca tradicionais. Acesse a Internet sem saber que existe uma falsa cortina usado pelo motor de busca mais convencional, é limitar as capacidades de navegar num oceano de informações valiosas, é aí que reside a importância de conhecer a web a partir da superfície (*Surface Web*) para deep (*Web Profundo*).

**Palavras-chave:** Superfície Web, TCP / IP, URL, WWW, motores de busca, HTML, Deep Web.

## Introducción

La Internet como se conoce hoy en día no surgió de la noche a la mañana. Entre los años cincuenta y sesenta del siglo XX, personajes como Paul Baran, Bob Kahn y Vinton Cerf (Ecured, 2013) abrieron el camino para consolidar una eficiente red de comunicación. Más tarde, Timothy John Berners-Lee

(NNDB, 2013) crearía un protocolo que permitió expandir los alcances de la Internet más allá de los laboratorios de universidades y oficinas militares. Con el surgimiento de la World Wide Web (Vences y Segura, 2011) se intentó mantener el orden amenazado por la entrada en masa de nuevos usuarios, quienes utilizaban lenguajes diferentes, no solo en sus idiomas, sino en sus máquinas y los formatos de su información. Años de trabajo, esfuerzo y gran inversión económica finalmente lograron convertir la Internet en el elemento fundamental en la cual se sostiene la información y la comunicación mundial. Esta particularidad provocó un evento imprevisible (aún en la actualidad es motivo de estudio, análisis y debate): los mecanismos de búsqueda utilizados por las personas no podían tener acceso a toda la información alojada en Internet, por lo que aparece un espacio en la red denominado *deep web* (Dragut, Meng y Yu, 2012; Liang, 2008), donde se encuentra ubicada toda la información que escapa de las herramientas de búsqueda comunes. Las propiedades de este espacio, unidas a las adquiridas a través de programas de anonimato, señalan una nueva dirección en el camino del desarrollo de la Internet y generan zonas ocultas en “un océano de información” del cual solo se conoce la superficie (Moreno, 2005).

Son innumerables las acciones, políticas y recursos destinados en el mundo para reducir la brecha digital y brindar la posibilidad de contar con acceso a Internet como un medio de comunicación de gran utilidad para la interacción y la investigación (Mochi, 2012), pues este recurso se ha convertido en la clave para la cotidianidad de las personas, gracias a la interconexión y la globalización de sistemas de cómputo, que permiten compartir y consultar información de una manera eficiente y rápida (Bernstein, Madhavan y Rahm, 2011). En la actualidad las personas alrededor del mundo hacen uso de esta herramienta, pero ¿realmente la sociedad sabe lo que es la Internet? ¿Acaso es consciente de las responsabilidades, riesgos y beneficios que se adquieren al ser parte de esta red mundial?

## Origen de la Internet

En el año 1958 el proyecto espacial Sputnik, desarrollado por la URSS durante la guerra fría, alarmó a los Estados Unidos, lo que llevó a la creación de la Advanced Research Projects Agency (ARPA), encargada del desarrollo de nuevas tecnologías de comunicación (Martínez, 2009).

Uno de los proyectos de esta agencia era solucionar los inconvenientes de seguridad que significaban el uso de una red que dependiera de un núcleo para las comunicaciones. Para resolverlo usaron una de las investigaciones realizadas por Paul Baran en la agencia RAND (Leiner et al., 1999).

La solución entregada por Baran fue implementada por primera vez en 1969: permitió conectar cuatro computadores ubicados en diferentes universidades en Estados Unidos (Utah, Stanford, Santa Bárbara y Los Ángeles), esta red fue llamada Arpanet y fue el desarrollo que abrió paso al nacimiento de la Internet (Moraga, 2003).

Con la implementación de este modelo, se envía el primer mensaje con la palabra "LOGWIN", el cual, a pesar de no transferirse completamente, sí demostró su funcionalidad (Martínez, 2009).

Cinco años después, los investigadores Bob Kahn y Vinton Cerf propusieron el protocolo TCP/IP, el cual fue definido como estándar años más tarde para la vinculación de cualquier máquina a la Arpanet (Estrada, 2004).

A mediados de 1983, gracias a la implementación del TCP/IP como estándar, Arpanet se separa de la red militar (Rodríguez, 2007), esto generó un aumento de usuarios en la red que reveló un inconveniente de compatibilidad en la transferencia de información, solucionado por *Tim Berners Lee* al crear la red mundial llamada *World Wide Web*, fundamentada en el Protocolo de Transferencia de Hipertexto o HTTP (McPherson, 2009).

Según datos de Red.es (2010), desde 1993 se han registrado más de 1.200.000 dominios, Las cifras globales apuntan a que en 2014 el número total de páginas web, bien de acceso directo mediante URL, bien mediante el acceso a través

de enlace, es de más de 27.000 millones (Lewandowski, 2014).

Diez años después de la primera página, la evolución de la WWW inició su fase 2.0, al aparecer la denominada *web social*. En 2004, Tim O'Reilly acuñaba el término de *web 2.0* o *web social* para referirse a una segunda generación de webs basadas en comunidades de usuarios y una gama especial de servicios que fomentan la colaboración y el intercambio ágil de información entre los usuarios. Trabajos como Allen, M. (2012) exploran la web 2.0, su forma de interactuar con el usuario y su evolución.

Esta evolución de la web convierte a los internautas en gestores de los contenidos con los que interactúan, de manera que pueden modificar utilidades y servicios de Internet que se sustentan en una base de datos, ya sea en su contenido o en la forma de presentarlos, o simultáneamente en contenido y forma (Rehman y Vaish, 2013).

Hoy en día, prácticamente todos los usuarios de Internet están acostumbrados a iniciar el explorador y entrar a algunos de los buscadores existentes en la red para buscar cualquier tipo de información (Rodríguez, 2007). Esta facilidad de búsqueda también genera debates éticos sobre la información que se busca y su uso adecuado, al igual del nivel comercial de todo tipo de información almacenada sin ninguna supervisión para que todo tipo de público la vea (Fernández, 2007).

Internet evoluciona a la WWW la cual es un espacio de búsqueda sencillo y difícil para usuarios inexpertos, de esta primera etapa se evoluciona a la web semántica, en donde se clasifica, almacena y concentra información relevante y veraz, esto ha llevado a que de la web inicial ya no queda nada (Berrío-Zapata, Vicentini y Gonçalves, 2014).

Con bastante certeza, el futuro de la web es la *web semántica*; sin embargo, algunos autores predicen que va hacia una red global del conocimiento (Chávez, Cárdenas y Benito, 2014).

Sin embargo, mucha gente desconoce que existe un mundo oculto fuera de la Internet conocido por todos, y mucho menos se sabe que este submundo

de Internet, al que es difícil llegar utilizando los medios tradicionales de conexión, incluye un mayor número de información del que utilizan la mayoría de los internautas (Krotoski, 2013).

Este lado oculto de Internet es el conocido como *internet invisible* (o internet profunda, deep web, surface web, entre tantos términos); está compuesto por la parte de Internet que no puede ser indexada por los motores de búsqueda (Ladrón de Guevara, Corradini, Bringa, Mansilla, Lamberti y Melián, 2012).

## Surface Web

Usualmente, cuando se habla de usar la Internet se expresa que se va a navegar en ella, y para hacerlo es necesario el uso de un navegador que permita acceder a las páginas de manera remota a través de su dirección URL (Vences y Segura, 2011). Esas páginas por lo general son de libre acceso y se encuentran en formato HTML, JSP, ASPX, PHP, entre otros, los cuales son interpretados por el navegador de modo que se brinda al usuario una interfaz amigable; este conjunto de páginas y servicios es lo que normalmente se conoce como la World Wide Web, la cual es una red del conjunto que conforman la Internet (Veiga, 2014). Esta es una red pública a la cual los usuarios acceden con poca o ninguna restricción, sin embargo, este nivel de libertad en el acceso es opacado por la falta de privacidad a la que están sometidos los usuarios, ya que es posible rastrear su ubicación y sus actividades a través de diferentes técnicas (Aguilar, 2011).

Actualmente, debido a la cantidad de sitios virtuales en los que se encuentra la información, es necesaria la ayuda de motores de búsqueda, los cuales, al escribir palabras claves sobre el tema de interés, arrojan un listado de direcciones en donde se encuentra información relacionada (Mampel y Gracia, 2009). Estas páginas de libre acceso son la respuesta de los motores de búsqueda a la consulta, y son lo que se conoce como *surface web*.

En un principio no existían los motores de búsqueda, las búsquedas se realizaban en índices temáticos que eran un listado de páginas agrupadas manualmente de acuerdo con su contenido (era como buscar en un directorio telefónico), este método hacía de la búsqueda un proceso demorado y limitado, debido a que era necesario revisar página por página para hallar la información requerida y estaba sometida a las “pocas páginas” que los desarrolladores hubieran tenido a bien incluir en estas listas (Castro, 2011).

En abril de 1994, David Filo y Jerry Yang, dos universitarios norteamericanos, crearon Yahoo! que albergaba una colección de las páginas web favoritas. El problema de Yahoo! es que comenzó siendo un directorio elaborado por personas, lo cual llevaba mucho tiempo por lo que tuvo que evolucionar, incorporándole un buscador (Parrilla 2012).

Posteriormente surge la primera generación de motores de búsqueda (*AltaVista, ASK, Lycos*, entre otros) que basaban su algoritmo en el análisis de tres factores del código HTML (título, descripción y palabras clave), pero dejó de ser funcional cuando los administradores de las páginas se dieron cuenta de que modificando estos atributos se podía perturbar la eficacia a la hora de mostrar los resultados de la búsqueda (Estrada, 2004). Debido a las falencias que se presentaban a la hora de realizar una búsqueda con estos algoritmos, se rediseñó su funcionamiento y surgió la segunda generación de motores de búsqueda, que es la actual y que funcionan basados en algunos programas: por un lado están los *spiders*, que son los encargados de recorrer la internet recopilando información sobre el contenido de las páginas para almacenarla en un índice; mientras que otro programa es el encargado de interactuar con el usuario para consultar la base de datos (Mampel y Gracia, 2009).

Para comprender cómo se navega por la *surface web* se tomará como ejemplo la forma en que se realizan las búsquedas a través del buscador Google:

este inicia su proceso dando un listado de URL de páginas web para ser visitadas por los *spiders* (SEO, 2012) quienes analizan los atributos usados por la primera generación de motores de búsqueda pasando de una página a otra a través de los enlaces indexados creando un índice con todas las páginas recorridas, para optimizar los resultados obtenidos en las búsquedas realizadas por el usuario, Google hace uso de un filtro que consta de más de 200 preguntas las cuales permiten organizar los resultados, se utiliza además el algoritmo creado por *Larry Page* y *Sergey Brin*, el cual calcula la relevancia de una página según el número de enlaces externos que la referencian (Brin, y Page 2012).

Todas las páginas indexadas que los *spiders* recorren y analizan son las que componen la *surface web*, estas páginas por lo general no están protegidas con ningún tipo de seguridad (contraseña, proxy, entre otros) o son de formatos estándar.

Empresas, organizaciones y cualquier persona que utiliza dispositivos de comunicación de datos, envía y recibe correos electrónicos, comparte información de manera local o a nivel mundial, realiza transacciones, ofrece servicios y encuentra soluciones a sus requerimientos requieren un nivel de seguridad en sus datos (Coello, 2013; Contreras y Bautista, 2012) y ahí es donde ya la *surface web* no ofrece lo requerido por el mundo digital que evoluciona constantemente.

Según Ferrara, De Meo, Fiumara y Baumgartner (2014), son dos las clases principales de uso de las aplicaciones en internet, a nivel de empresa y a nivel social. En el ámbito empresarial surgen técnicas de extracción de datos web como una herramienta clave para llevar a cabo el análisis de datos en los sistemas de inteligencia de negocios y competitivas (Malca, 2014). En cuanto al nivel social en Internet, las técnicas de extracción de datos web permiten reunir una gran cantidad de datos estructurados generados continuamente y difundidos por usuarios de redes sociales en línea de medios sociales.

El contenido web al que se puede acceder a través de hipervínculos se denomina *surface web*, mientras que los contenidos ocultos escondidos detrás de los formularios HTML se han denominado *deep web*. Trabajos como los de Sharma y Sharma (2011) se concentran en descubrir las ventajas y limitaciones de los actuales rastreadores en la búsqueda de los contenidos de la *deep web*.

## Deep Web

En la Internet existen muchos caminos para encontrar información, la dificultad se presenta cuando los buscadores no logran llegar a ese nivel donde la información se encuentra alojada. Fransson (2010) muestra un detallado análisis del concepto de búsqueda y las razones por las cuales los usuarios no pueden hallar la información relevante. Sin embargo, hay que resaltar que la mayoría de los usuarios de internet no requieren de búsquedas avanzadas, dado que la información solicitada es de ámbito social y empresarial, lo que lleva a reducir las búsquedas especializadas a una pequeña porción de la población (Pohorec, Zorman y Kokol, 2013).

Los usuarios recurren a los buscadores para tener acceso a la información, pero a la vez se preocupan por asegurar su información. Este fenómeno es muy común en las empresas y viene creciendo en las personas, a quienes les preocupa qué se hace con sus datos; por esta razón, la seguridad comercial y personal se incrementa cada día más y se ven iniciativas por asegurar lo que se envía a través de la red (Garfinkel y Spafford, 2002). Esto lleva a invisibilizar nuestros datos y ocultarlos de los buscadores tradicionales o generar información que no es real para desviar la atención de la información que se quiere cuidar (I Vilardebó, 2011).

Los motores de búsqueda generan gran cantidad de respuestas, lo que facilita acceder a la información, sin embargo, los buscadores convencionales tienen limitantes (a pesar del gran esfuerzo que hacen las compañías). Se calcula que los

motores de búsqueda solo indizan entre la mitad y un tercio de los documentos públicos de Internet (Sharma y Sharma, 2011), aun así, es una enorme cantidad de información. Se debe tener en cuenta que, por grande que parezca esta cantidad, no está ni cerca de comprender la totalidad de la información contenida en la web (Parrilla, 2012). El éxito de un buscador depende en gran medida de la cantidad de páginas que es capaz de recorrer y almacenar en sus bases de datos y de la forma de organizarlas para brindar una amplia gama de posibilidades con calidad a los usuarios (Gupta y Bhatia, 2012).

Existen redes en las cuales los buscadores no pueden indizar sus páginas, manuales o documentos, debido a su estructura o funcionamiento, por esta razón es necesario usar programas específicos. Según un estudio realizado por BrightP “Cursos de Introducción a Guadalinex”, la red es aproximadamente 550 veces mayor de la que se puede encontrar a través de los motores de búsqueda convencionales y crece con mayor velocidad (Cafarella, Halevy y Madhavan, 2011). Sin embargo, el método de estimación usado en este estudio no es el único existente (Liang, 2008).

Al igual que las redes, existen aplicaciones para cifrar información y llevarla al nivel profundo del “océano” de información. Las empresas son los principales usuarios de estas aplicaciones, al igual que existe todo un comercio ilegal de información, el cual no desea ser detectado por los buscadores tradicionales (Kaczmarek y Węckowski, 2013; Madani y Khalil, 2014). Esto ha generado un reto para los gobiernos y empresas dedicadas a resguardar la integridad de la Internet pero la batalla se está perdiendo (Luján, 2013), pues por cada portal encontrado se puede pensar que existen diez que no se encuentran.

Dentro del conjunto de elementos que componen la información de la *deep web*, Sherman y Price (2003) identificaron cuatro tipos de contenidos según una serie de características que demarcan la “profundidad” en la que se encuentran a la hora de acceder a ellos:

- Los archivos que podrían estar incluidos en los motores de búsqueda pero no se encuentran por razones como: profundidad de rastreo, esto quiere decir que no todas las páginas de un sitio son indizadas; frecuencia de rastreo, hace referencia a la frecuencia con que se actualizan las bases de indizaciones; número máximo de resultados visibles, es decir, los motores de búsqueda generalmente limitan el número de resultados (Unold, 2011).
- Los enlaces que pueden estar indizados, pero son excluidos por razones como el uso de métodos (“robots.txt” o campo “noindex”) para no ser indexados o con protección por contraseña.
- Contenido propietario, es necesario registrarse para tener acceso a él, ya sea de forma gratuita o paga.
- Páginas que no pueden ser indizadas por limitaciones técnicas de los buscadores o por sus características propias, por ejemplo, formatos diferentes al HTML, páginas generadas dinámicamente o información almacenada en bases de datos relacionales, que requieren de una búsqueda específica (Sherman Price, 2003).

Pero no hay que dejar a un lado el concepto de red que es la base de la Internet:

Las redes en general consisten en compartir recursos, y uno de sus objetivos es hacer que todos los programas, datos y equipos estén disponibles para cualquiera de la red que así lo solicite, sin importar la localización física del recurso y del usuario (Castro, 2011).

Esto lleva a pensar si un usuario no quiere ser visto, no puede ser visto su recurso, pero ahí es donde la indización se queda corta.

Sumado a la no indización por las características mencionadas, aparece otro factor que lleva esta red a un nuevo nivel. Se trata del uso de programas que además de facilitar el acceso a la *deep web* ofrecen herramientas que les permiten a los

usuarios aumentar su nivel de anonimato, lo cual les brinda la posibilidad de expresarse de forma abierta, dejando de lado algunas de las restricciones de tipo legal o social que controlan el uso de la *surface web* (Carrasco, 2012).

Algunas organizaciones han desarrollado, dependiendo de los intereses de los posibles usuarios, proyectos que buscan proveer anonimato a la hora de conectarse a una red, por ejemplo: TOR, i2p y FreeNet (Pasquier, 2010).

El problema principal que se presenta en la *deep web* es la consistencia de los datos. Li, Dong, Lyons, Meng y Srivastava (2012) encontraron alta inconsistencia en los datos de diferentes fuentes, además de fuentes con escasa precisión, lo que ocasiona poca veracidad en los datos obtenidos en esta porción significativa de la Internet. No obstante, como se dijo, muchos usuarios dejan información no real para distraer a los buscadores y desviar su accionar de los datos que poseen valor; de ahí se puede concluir inicialmente la importancia de encontrar la verdad en las búsquedas realizadas a este nivel en la internet (Saboori y Mohammadi, 2012).

Desarrollos como *OXPath* buscan solucionar precisamente el problema de las búsquedas avanzadas en este nivel de la Internet. Según sus autores (Furche, Gottlob, Grasso, Schallhart y Sellers, 2013), este desarrollo supera de forma sustancial a los buscadores comerciales y a algunos desarrollados para la búsqueda de información oculta; sus principios son claros: realizar una búsqueda escalada y con un tiempo determinado en cada uno de los niveles que se pueden explorar. En su trabajo manifiestan cuatro aspectos fundamentales para avanzar en la búsqueda: (1) interactuar con interfaces de aplicaciones web sofisticado, (2) capturar con precisión los datos pertinentes para ser extraídos, (3) escala con el número de páginas visitadas e (4) integrar fácilmente en tecnologías web existentes.

Es importante tener en cuenta que la búsqueda en *deep web* es un proceso de dos etapas de selección de las fuentes de alta calidad y la clasificación

de los resultados de las fuentes seleccionadas (Balakrishnan, Kambhampati y Jha, 2013). Aunque hay métodos existentes, tanto para los pasos, evalúan la pertinencia de las fuentes y los resultados utilizando la similitud de la consulta. Cuando se aplica a la Deep Web estos métodos tienen dos deficiencias: en primer lugar es agnóstico a la corrección (fiabilidad) de los resultados; en segundo lugar, la relevancia basada en consultas no considera la importancia de los resultados y las fuentes.

## Conclusiones

Si se analiza atentamente la historia, se puede decir sin lugar a duda que la Internet es la herramienta de comunicación más útil que ha desarrollado la humanidad. Desde sus orígenes ha servido de apoyo para solucionar los problemas de quienes han encontrado la forma de aprovechar al máximo su potencial. La masificación de su uso ha vinculado a la humanidad en el proceso evolutivo de la Internet, tanto así que se ha convertido en una virtualización del mundo real, con todo lo que esto implica.

Gracias a los esfuerzos realizados por las compañías encargadas del desarrollo y evolución de los motores de búsqueda, es posible indizar gran cantidad de archivos no HTML, lo que permite acceder a mayor información y mantener las bases de indización actualizadas para brindar mejores resultados a la hora de hacer una consulta en la red.

Los avances realizados por los motores de búsqueda sugieren que poco a poco la *deep web* irá disminuyendo su volumen, aunque la realidad es que se ha convertido en una combinación de privacidad y anonimato, características proporcionadas por algunas redes como TOR, las cuales han demostrado tener una aplicabilidad que permite trascender a un nuevo nivel el concepto de libertad a la hora de conectarse a Internet.

Existen realidades humanas que son plausibles y admirables, que surgen en procura del bienestar general, pero también existen realidades que

no encajan dentro de lo socialmente aceptable, así pues, la Internet seguirá evolucionando y brindando nuevas posibilidades que serán el reflejo de las intenciones con que los usuarios decidan utilizar esta herramienta.

## Referencias

- Aguilar, A. (2011). Ningún navegador es seguro. *Seguridad Defensa Digital*, 9.
- Allen, M. (2012). What was Web 2.0? Versions as the dominant mode of internet history. *New Media & Society*, 1461444812451567.
- Balakrishnan, R., Kambhampati, S., & Jha, M. (2013). Assessing relevance and trust of the deep web sources and results based on inter-source agreement. *ACM Transactions on the Web (TWEB)*, 7(2), 11.
- Baran, P. (1962). *On Distributed Communications Networks*. The RAND Corporation, Santa Monica, California. p. 2626.
- Bernstein, P. A., Madhavan, J., & Rahm, E. (2011). Generic schema matching, ten years later. *Proceedings of the VLDB Endowment*, 4(11), 695-701.
- Berrío-Zapata, C., Vicentini, M. J., & Gonçalves, R. C. (2014). Web technological path and digital order in Latin America: historical reflections from Brazil. *Revista Interamericana de Bibliotecología*, 37(2), 127-140.
- Brin, S. y Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 56(18), 3825-3833.
- Cafarella, M. J., Halevy, A. y Madhavan, J. (2011). Structured data on the web. *Communications of the ACM*, 54(2), 72-79.
- Carrasco, L. (2012). Redes de anonimización en internet: cómo funcionan y cuáles son sus límites. Instituto Español de Estudios Estratégicos.
- Contreras, J. y Bautista, D. (2012). Implementación de la seguridad del protocolo de internet versión 6. *Gerencia Tecnológica Informática*, 11(29), 35-46.
- Coello, L. (2013). Seguridad en redes.
- Dragut, E. C., Meng, W., & Yu, C. T. (2012). Deep Web Query Interface Understanding and Integration. *Synthesis Lectures on Data Management*, 7(1), 1-168.
- Chávez, M., Cárdenas, O. y Benito, O. (2014). La web semántica. *Revista de Investigación de Sistemas e Informática*, 2(3), 43-54.
- Ecured. (2013). *Baran, Kahn, Cerf*. [En línea]. Recuperado de <http://www.ecured.cu/index.php> Consultado el 27 de agosto del 2013.
- Estrada, A. (2004). Protocolos TCP/IP de Internet. Coordinación de Publicaciones Digitales. DGSCA-UNAM, 5(8).
- Castro, F. (2011). Redes de computadoras.
- Fernández, F. (2007). *Ética e Internet: manzanas y serpientes*. Ediciones Rialp.
- Ferrara, E., De Meo, P., Fiumara, G., & Baumgartner, R. (2014). Web data extraction, applications and techniques: A survey. *Knowledge-Based Systems*, 70, 301-323.
- Furche, T., Gottlob, G., Grasso, G., Schallhart, C., & Sellers, A. (2013). OXPath: A language for scalable data extraction, automation, and crawling on the deep web. *The VLDB Journal*, 22(1), 47-72.
- Fransson, J. (2010). *Efficient Information Searching on the Web*. Jonas Fransson.
- Garfinkel, S. y Spafford, G. (2002). *Web security, privacy & commerce*. "O'Reilly Media, Inc."
- Gupta, S., y Bhatia, K. (2012). Exploring 'hidden' parts of the web: the hidden web.
- I Vilardebó, L. F. (2011). Aspectos avanzados de seguridad en redes. Editorial UOC.
- Kaczmarek, T., y Węckowski, D. (2013). Harvesting Deep Web Data through Produser Involvement. *Frameworks of IT Prosumption for Business Development*, 200.
- Krotoski, A. (2013). *Untangling the web*. Faber & Faber.
- Ladrón de Guevara, M. C., Corradini, L., Bringa, V., Mansilla, G., Lamberti, A. y Melián, J. (2012). Los docentes investigadores de la Facultad de Filosofía y Humanidades de la Universidad



- Nacional de Córdoba como usuarios remotos de bibliotecas universitarias. *Información, Cultura y Sociedad*, 26, 65-81.
- Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., ... & Wolff, S. (1999). Una breve historia de Internet. Recuperado de <http://www.ati.es/DOCS/internet/histint/histint1.html>.
- Liang, J. (2008). Estimation methods for the size of deep web textual data source: A survey. Recuperado de [cs.uwindsor.ca/richard/cs510/survey/jie-liang.pdf](http://cs.uwindsor.ca/richard/cs510/survey/jie-liang.pdf).
- Li, X., Dong, X. L., Lyons, K., Meng, W., & Srivastava, D. (2012). Truth finding on the deep web: is the problem solved? *Proceedings of the VLDB Endowment*, 6(2), 97-108.
- Lewandowski, D. (2014). Evaluating the retrieval effectiveness of Web search engines using a representative query sample. [En línea]. Recuperado de [arXiv:1405.2210](http://arxiv.org/abs/1405.2210).
- Luján, S. (2013). Web Accessibility Among the Countries of the European Union: a Comparative Study.
- Malca, Ó. (2014). *Comercio electrónico*. Universidad del Pacífico, Centro de Investigación.
- Mampel, L., y Gracia, Á. (2009). Los motores de búsqueda de información en Internet como herramienta de aprendizaje: análisis de contenidos sobre el concepto "dinosaurio". *REEC: Revista Electrónica de Enseñanza de las Ciencias*, 8(3), 2.
- Madani, S., y Khalil, I. (2014). Garbled Routing (GR): A generic framework towards unification of anonymous communication systems. *Journal of Network and Computer Applications*, 44, 183-195.
- Martínez, C. (2009). Bases tecnológicas de la telemedicina. (III) . Internet. Origen y Evolución. REDUCA, 1(1) .
- Mochi, P. (2012). Programas para la inclusión digital y la concertación de actores en procesos de desarrollo territorial. *Polis*, 8(1), 177-212.
- Moraga, A. (2003). Historia e Internet: aproximación al futuro de la labor investigadora. En *Aportacions de la comunicació a la comprensió i construcció de la història del segle XX; La comunicació audiovisual en la història* (pp. 369-395). Servicio de Publicaciones.
- Moreno, P. (2005). Estrategias y mecanismos de búsqueda en la web invisible.
- McPherson, S. (2009). *Tim Berners-Lee: Inventor of the World Wide Web*. Twenty-First Century Books.
- NNDB. (2013). Tim Berners-Lee. [En línea]. Recuperado de <http://www.nndb.com/people/573/000023504/>. Consultado el 29 de agosto del 2013.
- Parrilla, M. (2012). La Internet que no aparece en los buscadores.
- Pasquier, M. (2010). Redes de Anonimato TOR. R. M. Stallman (comp.) *Directorio de software libre con documentación libre y biblioteca de libros libres*.
- Pohorec, S., Zorman, M., y Kokol, P. (2013). Analysis of approaches to structured data on the web. *Computer Standards & Interfaces*, 36(1), 256-262.
- Rehman, V., & Vaish, A. (2013). Content fusion in traditional and new media: A conceptual study. *Journal of Internet Commerce*, 12(2), 225-245.
- Rodríguez, A. (2007). Iniciación a la red Internet. Concepto, funcionamiento, servicios y aplicaciones de Internet. Ideaspropias Editorial.
- Saboori, E. y Mohammadi, S. (2012). Anonymous communication in peer-to-peer networks for providing more privacy and security. *arXiv preprint arXiv:1208.3192*.
- SEO, A. (2012). Google: Cómo funciona su algoritmo de búsqueda.
- Sharma, D. K., & Sharma, A. K. (2011). A novel architecture for deep web crawler. *International Journal of Information Technology and Web Engineering (IJITWE)*, 6(1), 25-48.
- Sherman, C., y Price, G. (2003). The invisible web: uncovering sources search engines can't see. *Library Trends*, 52(2), 282-298.
- Unold, J. (2011). The invisible web in enhancing managerial creativity. *Support Systems*, 169.

Veiga, A. (2014). Query scheduling techniques and power/latency trade-off model for large-scale search engines (Doctoral dissertation). Departamento de Tecnoloxías da Información e as Comunicaciós, Universidade da Coruña.

Vences, N. A., & Segura, R. V. (2011). El desarrollo de la World Wide Web en España: Una aproximación teórica desde sus orígenes hasta su transformación en un medio semántico. *Razón y Palabra*, 75 (64).

